

P2P-VoD 系统中自适应大小的滑动窗口模型研究

黄益贵, 王汝传

(南京邮电大学 计算机学院, 江苏 南京 210003)

摘要: VoD 业务是当今互联网上广泛流行的业务, 人们通过在线浏览自己喜爱的新闻、娱乐、电影等视频剪辑, 但在互联网上开展 VoD 业务仍是一项艰巨的挑战, 这主要是因为海量的媒体数据存储和高额的带宽消耗以及需要高性能的服务运算能力所带来的巨额开销。P2P-VoD 的出现无疑给人们带来了从根本上解决问题的希望, 它通过将数据分储在每个参与的节点上, 对每个业务的使用者通过一些策略使其同时成为业务的提供者。现在的 Internet 上已经部署了一些可以使用的 P2P-VoD 系统, 但是这些系统仍然普遍存在播放启动时延较大和系统负载均衡较差等诸多问题。文中提出的方案试图解决和优化以上遇到的问题。试验结果数据表明, 该方案能够有效改善播放的启动时延(减少为原来的 1/2)和具有良好的系统负载均衡效果。

关键词: P2P-VoD; 滑动窗口; 播放启动时延; 负载均衡

中图分类号: TN919.85

文献标识码: A

文章编号: 1673-629X(2010)05-0021-05

Research of Self-Adjust Size of Sliding Window Model in P2P-Based VoD System

HUANG Yi-gui, WANG Ru-chuan

(College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Nowadays, the VoD service is widely popular on the Internet, people glance over video clips of news, entertainment, film on the line. It is still an arduous challenge to deploy VoD service on the Internet, since the magnanimous media data storage and the high quota band width consumed as well as the large amount expenses needed which the high performance the service operational capability brought. No doubt that P2P-VoD's appearance would bring hope that fundamentally solved the question, movies are split into chunks that can be stored at partners, each service's consumer is the service's provider simultaneously by some strategies. On present's Internet had already deployed some P2P-VoD systems, but these systems still generally had seriously playback startup latency and low load balancing, etc. The paper proposed attempts to solve and optimize the meets. The results show that our mechanism is effective in suppressing the playback startup latency(drops to original 1/2) and reducing the workload of P2P-VoD system.

Key words: P2P-VoD; Sliding Window; playback startup latency; load balancing

0 引言

随着互联网宽带技术和音视频编解码技术的发展, VoD 成为互联网上方兴未艾的技术, 但是传统的基于一对多的 C/S 架构存在着服务用户数量有限、服

务器出口带宽成为 VoD 系统的“瓶颈”、海量数据存储在服务器端开销过大等诸多问题。P2P-VoD 技术的出现无疑能从根本上解决以上问题, 因而关于 P2P-VoD 的研究也成为备受瞩目的课题, 现在已经有一些相对成熟的 P2P-VoD 系统部署在互联网上, 提供 P2P 点播服务。如广为熟知的 PPLive、PPStream。这些系统初步实现了基于 P2P 的 VoD 服务, 给人们在线点播带来了全新的感受, 改变了过去人们需要使用 BitTorrent 和 Emule 等 P2P 文件下载工具“先下载、后观看”的浏览模式。以上系统虽然通过试验或仿真表明其在一定程度上实现了 P2P-VoD 服务, 但是以上系统或方案仍然存在如下问题: 1) 延迟较大, 从用户选定一个节目到节目开始播放, 等待的时间过长; 2) 资源调度和负载均衡, 在相同的网络环境下, 当用户数量达

收稿日期: 2009-07-23; 修回日期: 2009-10-11

基金项目: 国家自然科学基金(60973139, 60773041); 江苏省自然科学基金(BK2008451); 国家高科技 863 项目(2007AA01Z404, 2007AA01Z478); 省级现代服务业发展专项资金; 南京市高科技项目(2007 软资 127); 江苏高校科技创新计划项目(CX08B-085Z, CX08B-086Z); 江苏省六大高峰人才项目(2008118)

作者简介: 黄益贵(1984-), 男, 安徽寿县人, 硕士研究生, 研究方向为计算机网络、对等计算和信息安全等; 王汝传, 教授, 博士生导师, 研究方向为计算机软件、计算机网络和网络、对等计算、信息安全、无线传感器网络、移动代理和虚拟现实技术等。

到一定的值后, P2P-VoD 系统的性能出现急剧下降的现象。

文中通过对 P2P-VoD 的分块选取机制进行深入研究, 提出改进 Peer 节点选择和分块获取的算法, 解决和优化 P2P-VoD 系统的播放启动延迟过大问题; 提出自适应大小的滑动窗口模型改善文献[1]提出的基于统计特征得出的静态滑动窗口模型当网络的访问量增大到一定大小时, P2P-VoD 系统负载显著增大, 性能急剧下降等问题。通过将文中所提出的方案应用到作者所开发的 IPTV P2P-VoD 系统中的试验数据分析, 可以有效改善播放启动延迟和系统的负载均衡等问题。

1 系统架构

目前 P2P-VoD 系统的结构大致继承了文件共享 P2P 的结构, 但是由于 P2P-VoD 系统大多要求内容具有可管理性、鲁棒性、稳定性等文件共享 P2P 结构忽略或不强调的特性, 因此又具有其特异性, 如 Gnutstream^[2]采用基于 Gnutella 协议的拓扑结构, 则是一种改进型集中目录式结构; 对于大多数 P2P-VoD 系统, 则采用的是一种混合式的拓扑结构, 如 PPLive、PP-Stream, 继承和结合了 C/S 和 P2P 架构的各自优势, 考虑到 IPTV P2P-VoD 系统具有的特性和综合衡量各种架构的优缺点, 文中基于 BitTorrent 协议, 继承与改进文献[1, 3, 4]所采用的 P2P-VoD 系统架构。

本系统框架如图 1 所示, Peer 节点按照先后加入的顺序, 从 Tracker 上获取可以利用的 Peer 节点列表, 然后将这些节点拥有的分块位图和自己拥有的分块位图匹配, 按照一定的算法规则向拥有所需要分块的节点请求分块, 同时周期性地向 Tracker 报告自己拥有

的分块信息, 为其它 Peer 节点提供内容分发服务。

2 自适应大小的滑动窗口模型

P2P-VoD 不同于文件下载, 传统的 BT 协议基于“最少者优先”(rarest-first)^[5]策略, 实现一种无序的下载方式, 而 VoD 服务显然不能够完全采取这种方式, VoD 必须要求播放器优先得到即将播放的媒体数据, 而对其它非紧急数据, 则可以从负载均衡的方面考虑^[6]。采取一种文件片段共享机制, PPLive^[7]采取的是“拖拉”策略(pull method), 即是优先采取“顺序下载”策略, 然后是“最少者优先”策略, 这种方案类似于滑动窗口模型, 文献[4]则提出了滑动窗口模型, 对滑动窗口模型给予了较为详尽的讨论。文中在已有的工作基础之上, 同样基于滑动窗口模型详细讨论怎样改善和优化, 使得进一步降低点播服务的延迟等待时间, 改善 P2P-VoD 系统资源调度和负载均衡等问题。

2.1 分块获取策略

P2P-VoD 系统中, 传输策略主要考虑的因素有两点, 即是: a) 最大化下载速率; b) 最小化系统负载。

有三种方案实现 Peer 节点的分块请求:

1) 为了快速得到所需要的某个分块, 从 Tracker 响应的 Peer 节点集中根据分块位图信息匹配选择一组节点, 发送该分块的请求, 这种方案能够有效地保证所请求的分块得到及时下载, 但是这种方案却增加了系统的负载开销。

2) 对播放器所需要的不同分块, 从 Tracker 响应的 Peer 节点集中根据分块位图信息匹配选择一组节点, 将这些分块请求按照某种策略同时发给这组节点中的各个节点, 一旦超时, 则将超时分块请求重定向到节点集中的其它节点, 这种方案虽然比方案 1) 取得所

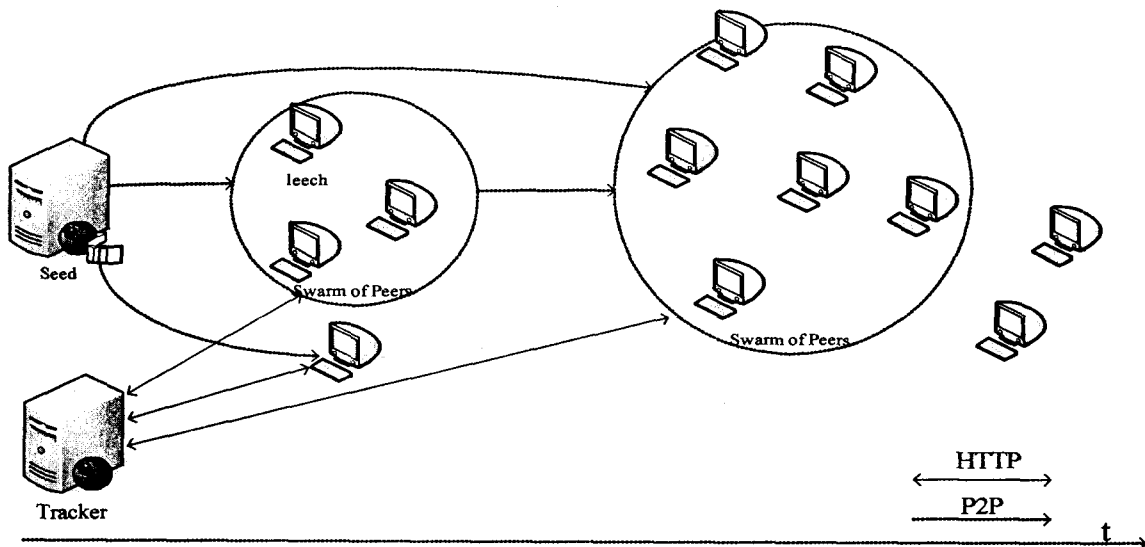


图 1 P2P-VoD 系统覆盖图

需要的片段花费的时间长,但是能够有效地减少系统的负载开销,提高系统中分块复制的利用率。

3) 对一个 Peer 请求,从节点集中选择一个节点,服务该 Peer 请求,如果该节点服务超时(比如离开该节点集,或网络延迟过大等),则从该节点集中选择另一个节点,这种方案对服务节点的服务能力要求比较高,对于中国大部分家庭的 ADSL 用户来说,难以负载。

我们在所设计的 P2P-VoD 网络中,采取了方案 1) 和方案 2) 结合的方法。对于滑动窗口内的播放器播放需要的紧急数据,采取方案 1), 按照紧急数据的优先级顺序,从节点集中选择一组节点,向该组节点发送该紧急数据的请求,这样能够保证紧急数据的快速得到,而对于位于滑动窗口外的非紧急数据,采取方案 2), 将这些分块请求分别发送给节点集中的不同节点,用于有效地改善系统负载。

2.2 伙伴节点选择策略

前面的讨论中阐述了对于滑动窗口内的分块请求和滑动窗口外的分块请求所采取的不同策略,在本小节中将详细阐述请求数据的 Peer 节点怎样对 Tracker 服务器响应的 Peer 节点集进行分类,以对资源构建合理的逻辑覆盖网,达到更加快速地获取所需要的分块。把从 Tracker 服务器得到的 Peer 节点分成两类: a) 高速节点; b) 低速节点。知道 Peer 节点得到所请求分块花费的时间,可分为两个部分,第一部分是 UDP 数据报传输时间,第二部分是响应节点进行响应的的时间,相对于高速传输的网络,响应时间完全可以忽略不计。因此给出如下定义:

假设 Peer 从 Tracker 上获取的节点集中共有 m 个可以提供下载服务,则当前能够提供下载服务的 Peer 节点集定义为: $\text{SetofServePeer} = \{p_1, p_2, p_3, \dots, p_m\}$ (其中,从 p_i 节点得到分块的速度定义为 b_i , b_i 即是 UDP 数据报传输时间)。则 Peer 节点集所能提供的平均下载速度定义为: $\text{Ave.} = \sum_{i=1}^m b_i / m$; 因此,可以很容易地定义提供高速下载服务的节点: $\text{HighSpeedPeer} = b_i (b_i \geq \text{Ave.})$; 提供低速下载服务的节点: $\text{LowSpeedPeer} = b_i (b_i < \text{Ave.})$; 提供高速下载的节点集所能提供的下载速度之和,也可以很容易计算出来: $\text{TSofHighSpeedPeers} = \sum b_i (i \in \text{HighSpeedPeer})$; 同理,提供低速下载的节点集所能提供的下载速度之和: $\text{TSofLowSpeedPeers} = \sum b_i (i \in \text{LowSpeedPeer})$ 。将滑动窗口内的紧急数据请求发送给能够提供高速下载的节点集,而将滑动窗口外的非紧急数据请求发送给提供低速下载的节点集,在实际的

试验中,按照滑动窗口内的分块优先级,将滑动窗口内的每个分块请求同时向 2~4 个节点请求为最佳,当然不同的分块可向同一个节点发送请求,只要保证高优先级的分块最先被请求,根据下面对分块上载部分的讨论,最先被请求的高优先级分块将最先得到响应,这样便能进一步减少分块请求的等待时间。

2.3 自适应大小的滑动窗口算法

滑动窗口维护一个关于媒体数据接收缓冲的窗口,这个窗口将根据播放器消耗数据和 Peer 客户端接收数据的状态变化向前移动,一旦位于滑动窗口内的数据被播放器消耗掉,则播放器将向前移动,以获取播放器后续播放需要的媒体数据。位于滑动窗口内的数据将根据紧急程度按照优先级顺序地从服务器或其它伙伴节点请求^[8]。

从前面的讨论可以知道,滑动窗口越大,则获取分块的速度越快,同时由于滑动窗口内每个分块都向多个 Peer 节点请求(以达到和保证迅速获取分块),无疑将增加系统中分块请求和响应的冗余,而滑动窗口越小,相对地,位于滑动窗口外的数据请求比率将会增大。对于那些位于滑动窗口外的数据,除了后面所采取的设置“锚点”的部分,仍然采用的是 BitTorrent 的“最少者优先”策略,系统的冗余将相对变小,能够有效地降低系统负载,但是这样可能不能够及时地满足播放器对即将播放数据的需求。这主要取决于播放器缓冲数据接收和消耗的速度的变化,当接收速度大于消耗速度的时候,显然缓冲能够满足播放器消耗的速度,这时候应该将滑动窗口减小,以降低系统负载,而滑动窗口将用于保证“最少者优先”策略没有下载下来的播放器所需的紧急数据的请求(“最少者优先”为了平衡系统负载,总是最先下载群集中复制较少的分块,这种无序下载的方式不能够满足播放器顺序播放的需要),当接收速度小于消耗速度的时候,因为播放器缓冲数据消耗较快,为了减少用户播放等待的时间,应该增大滑动窗口,以尽快获取播放器所需要的数据,因此,为了动态地反映滑动窗口内数据的变化,以达到更好地平衡系统负载,提出一种自适应大小的滑动窗口方案。

从上述论述可以清楚地认识到,滑动窗口的大小是一个影响 P2P-VoD 系统性能的至关重要的参数。我们认为滑动窗口的大小应该与播放启动等待时间 d 相对应, P2P-VoD 客户端应该在这个播放等待时间内填满滑动窗口缓冲,而后开始播放媒体流,在播放的同时,不断地继续填充这个缓冲,以保证播放的连续性。这样关于滑动窗口的最佳大小,文献[9]提出,根据试验统计,滑动窗口的大小应该静态地设置为能够缓冲 300s 的媒体数据。而文献[1]给出 dp/c (其中 d :

播放器的播放延迟,即从客户端请求节目到播放器开始播放之间的时间间隔; c :分块的大小,用于保存在 Peer 客户端的媒体数据的最小单位。大小为 256k 字节; p :媒体文件的解码速度)的滑动窗口大小计算公式,这两种方案本质上是一样的,他们方案的前提都在于假设分块获取的速度总是至少大于或等于播放器流媒体的解码速度,这在实际应用中,特别对于广大的 ADSL 用户和具有高质量 QoS 保证的流媒体数据,具有极大的不现实性。我们滑动窗口的初始大小同样采取文献[4]的计算方案,同时鉴于 P2P-VoD 分块获取的波动性,以及前述所讨论的滑动窗口对系统整体性能的影响,我们的自适应大小滑动窗口将根据分块获取的速度动态地自我调整大小。

首先给出如下定义:

s :节点 p_i 下载分块的速度之和,是实时动态的反映,根据上面的讨论,可得知 s 的大小为: $s = Ave. * m = \sum_{i=1}^m b_i$; ISWS:初始滑动窗口大小(理想滑动窗口大小),大小可由如下公式计算:ISWS = dp/c (其中,SSWS $_t$: t 时刻的自适应滑动窗口大小),则定义 SSWS $_t$ 的计算公式如下:

$\forall i, i \geq 1, i \in N^+$, 当 $\nabla t > t$ 时,有如下关系:

$$SSWS_i = \begin{cases} SSWS_{i-1}(1-k), & \text{当 } s > p \text{ 时} \\ SSWS_{i-1}, & \text{当 } s = p \text{ 时} \\ SSWS_{i-1}(1+k), & \text{当 } s < p \text{ 时} \end{cases}$$

其中 N 是节点所请求文件的分块数, $\nabla t = t_i - t_{i-1}$, t 是调整滑动窗口大小的频度。关于滑动窗口的最大值和最小值: (N, n) , 由前面知道 N 为分块数, 假设一个 512MB 的视频文件, 则 $N = 512 * 1024 * 1024 / (256 * 1024) = 2048$, 而 n 为滑动窗口的最小值, 理论上, 按顺序下载的话, 当分块获取速度大

于缓冲区消耗速度时, n 可以趋近于 0, 即是不需要滑动窗口的存在, 但是正如前面所讨论的, 由于 BT 无序下载的特性, 必须保留滑动窗口为一定的大小, 通过试验, 发现当解码速度为 4Mbps, 播放延迟为 5s 时, n 取值为 10 ~ 20 为最佳, 而 k 取值 15% ~ 20% 为最佳。

3 方案效果评估

试验评估数据来自于我们开发的 IPTV P2P-VoD 系统, 这套系统部署有 30 套 TI 公司基于达芬奇技术的 TMS320DM6446 数字媒体开发平台, 该开发平台具有 ARM 和 DSP 双核 CPU, 其中 ARM 用于资源处理运算, 而 DSP 则专司音视频解码, ARM 为 ARM9, 具有 283MHz 的主频, 拥有一个 100Mbps 的以太网接口和一块 80G 的硬盘。而 Seed 部署在一台 AMD Athlon 64X2 双核 4400+ 主机上, 该服务器的单个 CPU 主频为 2311MHz, 内存为 2G, 红帽子企业版 Linux 操作系统, 内核版本 2.6.9-42, 有一个 1000Mbps 的以太网接口和一块 250G 的硬盘。

图 2 左边反映参与节点数与片段获取之间的关系, 从图中可以看到当参与节点数在 5 个以上, 片段获取的成功概率为 92% 以上; 图 2 右边反映分块获取的分布关系, 从图中发现当节点刚加入时候, 大多从服务器上获取分块, 而随着时间的增加, 从 Peer 节点获取分块的比率逐渐增加, 并趋于一个稳定值。图 2 反映了系统的运行状况和负载均衡, 从这两个图中, 可以发现文中所提出的方案在一定程度上, 改善了系统的负载均衡, 具有良好的数据调度机制。

图 3 对比了文献[1]所描述的系统的播放启动时延和经过本方案优化后的系统的播放启动时延, 从图中可以看到, 原方案的播放启动时延大约在 10s 左右,

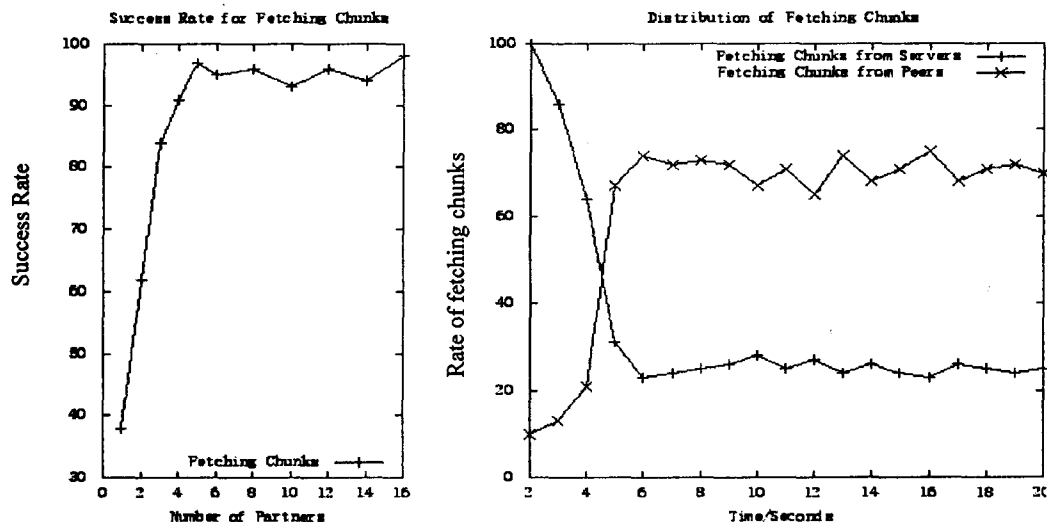


图 2 系统运行和负载均衡

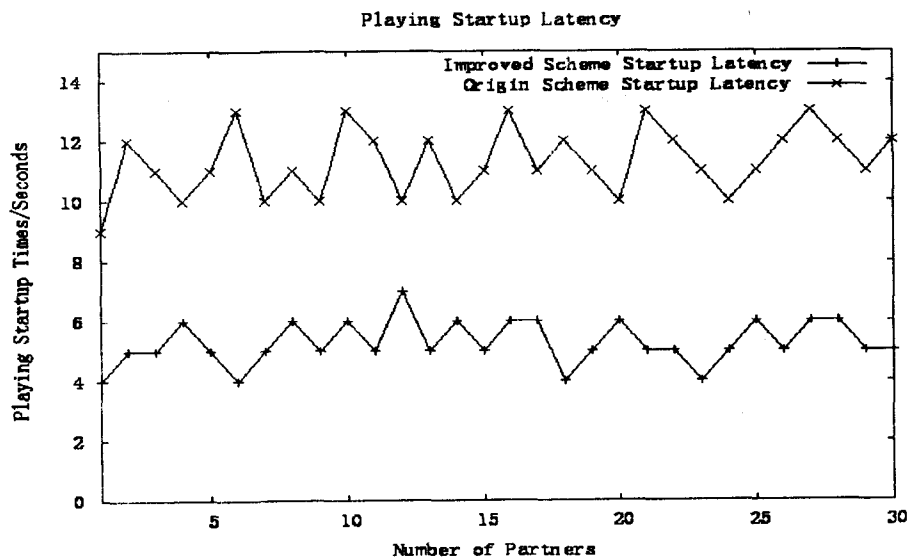


图3 改进方案的播放启动时延和原方案的播放启动时延对比关系

而改进后方案的播放启动时延大约在5s左右,是原来的1/2,这有效地改善了用户体验感受。

4 结束语

通过对基于BitTorrent协议的改进,提出了基于自适应大小的滑动窗口模型解决和优化P2P-VoD系统部署过程中所遇到的启动时延较长和系统负载均衡较差等问题。试验结果表明所提出的伙伴节点选择策略、分块获取策略、分块服务策略等方案能够在一定程度上解决所提出的问题。当然P2P-VoD系统的应用和完善还有诸如版权控制、系统冗余控制、带宽管理、QoS保证、可扩展性、可靠性、鲁棒性等诸多问题,将对这些方面做进一步的研究。

参考文献:

- [1] Cheng B, Liu X, Zhang Z, et al. A Measurement Study of a Peer-to-Peer Video-on-Demand System[M]//IPTPS. Bellevue, WA:[s.n.],2007.
- [2] Jiang X, Dong Y, Xu D, et al. GnuStream: a P2P Media

streaming system prototype[C]//In Proceedings of the 4th International Conference on Multimedia and Expo. Baltimore, Maryland:[s.n.],2003.

- [3] Luo J, Zhang Q, Tang Y, et al. A Trace-Driven Approach to Evaluate the Scalability of P2P-Based Video-on-Demand Service[J]. IEEE Transactions on Parallel and Distributed Systems, 2009,20(1):59-70.
- [4] Shah P, Páris J-F. Peer-to-Peer Multimedia Streaming Using BitTorrent[C]//In IPC-CC 2007. New Orleans, USA:[s.n.],2007.

- [5] Cohen B. Incentives build robustness in BitTorrent[C]//In Proc. of First Workshop on Economics of Peer-to-Peer Systems. Berkeley, CA:[s.n.], 2003.
- [6] Lu Z, Zhang S, Wu J, et al. Design and Implementation of a Novel P2P-Based VOD System Using Media File Segments Selecting Algorithm[C]//In 7th IEEE Intern. Conf. on Computer and Information Technology (CIT 2007). Washington DC, USA: IEEE Computer Society,2007:599-604.
- [7] Huang Y, Fu T T J, Chiu D M, et al. Challenges Design and Analysis of a Large-scale P2P VoD System[C]//In Proceedings of ACM SIGCOMM 2008. Seattle, Washington, USA:[s.n.], 2008.
- [8] Cui Y, Li B, Nahrstedt K. oStream: Asynchronous Streaming Multicast in Application-layer Overlay Networks[J]. IEEE Journal on Selected Areas in Communications. Special Issue on Recent Advances in Service Overlays,2004,22(1):91-106.
- [9] Janardhan V, Schulzrinne H. Peer assisted VoD for set-top box based IP network[C]//In Workshop of Proc. of ACM SIGCOMM,P2P-TV'07. Kyoto,Japan:[s.n.], 2007.

(上接第20页)

- System for Financial Forecasting[J]. Pattern Analysis and Applications,1999,2(3):264-273.
- [10] Zhang G P. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model[J]. Neuron-computing,2003,50(1):185-198.
- [11] 张立明. 人工神经网络的模型及其应用[M]. 上海:复旦大学出版社,1993.
- [12] Leigh W, Hightower R, Modani N. Forecasting the New York Stock exchange composite index with past price and invest rate on condition of volume spike[C]//Expert System with Appli-

cations,2005. Cambridge:Cambridge University Press,2005.

- [13] Yam J Y F, Chow T W S. Feed forward networks training speed enhancement by optimal initialization of the synaptic coefficients[J]. Neural Networks, IEEE Transactions, 2001, 34(5):73-85.
- [14] Alan M S. The application of neural networks to predict abnormal stock returns using insider trading data[C]//Applied Stochastic Models in Business and Industry, 2002. [s.l.]: MIT Press,2002.